DOCUMENT RESUME

ED 079 338 TM 002 936

AUTHOR Centra, John A.
TITLE The Student as Godfather? The Impact of Student Ratings on Academia.
INSTITUTION Educational Testing Service, Princeton, N.J.
REPORT NO ETS-RM-73-8
PUB DATE May 73
NOTE 19p.; Paper presented at the Invitational Conference on Faculty Effectiveness as Measured by Students (1st, Temple University, April, 1973)

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *College Students; *Curriculum Evaluation; Higher Education; Speeches; *Student Attitudes; *Student College Relationship; *Teacher Evaluation

ABSTRACT

The impact or possible impact of college student ratings on the individual instructor, on teaching generally, on students, on administrators, and on the college is discussed. A study of over 400 faculty members in which half were assigned to an experimental group and half were controls, showed that as a result of student ratings on an instructor's practices, changes in instruction occurred after only a half semester for instructors who were "unrealistic" in how they viewed their teaching, and a wider variety of instructors changed if given more than a half semester and if they were given minimal information to help them interpret their scores. Some adverse effects of student ratings are that the ratings do not allow for individual styles of teaching, and they encourage traditional modes of teaching. Flexibility in the employment of student ratings is extremely critical. Student ratings influence college administrators in that these evaluations make the administrator's job easier and more effective. Student evaluations may be contributing to the current interest in administrator evaluations by faculty members. Where student ratings have been incorporated into faculty evaluation procedures, the impact on students is likely to be positive. Probably the major impact of student ratings on students is provided by published course and teacher critiques. A worthwhile use of student ratings is that of providing departments with information about the effectiveness of their offerings as seen by students. Focusing on weaknesses highlighted by student evaluations could be applied at the college level. (DB)

ED 079338

TM 002 936

RESEARCH MEMORANDUM

RM-73-8

THE STUDENT AS GODFATHER?

THE IMPACT OF STUDENT RATINGS CN ACADEMIA

John A. Centra

Paper presented at the First Invitational Conference on Faculty Effectiveness as Measured by Students, Temple University, April 1973.

Educational Testing Service
Princeton, New Jersey
May 1973

FILMED FROM BEST AVAILABLE COPY

The Student as Godfather?

The Impact of Student Ratings on Academia[1]

John A. Centra

Most of you, I'm sure, are familiar with the Godfather role made popular by the very successful book and movie. He was depicted as someone with a great deal of power over people and viewed by most with a mixture of awe, fear, and respect. In fact, his "offers that one could not refuse" were indeed, as some of you will recall, quite compelling.

There are some who fear that the college student, by virtue of the apparent increasing emphasis on student ratings of professors, could become the "Godfather" of the academic community. More exactly, they fear that too much emphasis could be put on these ratings and that, generally speaking, the power that students might acquire would not be in the best interest of the academic community.

These Cassandras can, in fact, point to the medieval universities as an example of unreasonable student influence over teachers. As Hastings Rashdall tells us in his writings about the medieval European universities, students at the University of Bologna not only paid teachers a "collecta" or fee (which apparently was determined by a teacher's ability to haggle), but they also could report teacher irregularities to the rector. For example, law texts were divided into segments, and each instructor was required to cover a particular segment by a specified date; to enforce

---

[1]Paper presented at the First Invitational Conference on Faculty Effectiveness as Measured by Students, Temple University, April 1973.

this statute, the rector appointed a committee of students to report on dilatory professors, who were then required to pay a fine for each day that they had fallen behind.

While few people would take seriously the possibility that students are on the verge of assuming the role they played in medieval days, some do question the ultimate impact of student evaluations on teaching and learning. I will be more specific about some of their reservations later in this paper. In addition, I plan to discuss evidence of the positive effects of student ratings, and finally, since the impact of student ratings on certain aspects of academic life is not totally known, I will speculate about some possible consequences.

I've grouped my comments within five categories and will discuss the impact or possible impact of student ratings on the individual instructor, on teaching generally, on students, on administrators, and on the college.

The Individual Instructor

First, let me begin by discussing the person the ratings are meant to influence most: the individual teacher. There has been a good deal of skepticism over how much effect the ratings actually have on changing or improving instruction--particularly when the results are seen only by the individual teacher. Faculty conservatism, when it comes to educational changes, has been a well-known tendency, although there are signs that it may be less true now than in the past. For example, I recently had occasion to look at the responses of some 2800 college teachers to the question, "When did you last make changes in the teaching methods you are using?" About a fourth indicated that they had never made changes. On the other hand, about half said that they had changed their methods during the past

two years. So it looks as if we should not indict all college teachers with the time-worn stereotypes of stodginess and traditionalism. Many apparently are willing to change their methods.

The question, though, is what causes teachers to change and, more germane to my topic, can ratings by students lead to any noticeable changes among college teachers? While a few investigators have noted that the ratings that teachers receive seem to improve over time, we know that we cannot assume a cause and effect relationship. Those changes could have been caused by any number of factors other than the initial student feedback.

One of the best ways to investigate the effects of student ratings on an instructor's practices is to employ an experimental design in which random groups of teachers receive feedback from students while other teachers--those in the control groups--do not. As some of you know I completed such a study within the past year with the cooperation of over 400 faculty members at five colleges. The details of that study are presented elsewhere (Centra, 1972), so I won't take the time to repeat them. But I would like to discuss briefly the results. The major conclusions of the study were, first, that changes in instruction (as assessed by repeated student ratings) occurred after only a half semester for instructors whose self-evaluations were considerably better than were their student ratings. If, in other words, teachers were especially "unrealistic" in how they viewed their teaching--unrealistic relative to their students' views, that is--then they tended to make some changes in their instructional practices, even though they had only a half semester to do so. I might add that such variables as the subject area of the course, sex of the instructor, and number of years the instructor had taught did not distinguish which

instructors made changes; or to put it another way, none of the subgroups of teachers formed by these variables were more likely to change. The second conclusion was that a wider variety of instructors changed if given more than a half semester of time and if they had some minimal information to help them interpret their scores. Let's consider briefly the implications of each of these findings.

Starting with the first result, why do you suppose changes in teaching procedures were related to the discrepancy between self-evaluations and student ratings? Actually this result was predicted at the outset of the study because there was fairly good reason to expect it, based on social psychological theory. As a matter of fact there are several similar theories that help explain the finding. Most are referred to as self-consistency or equilibrium theories, the central notion being that an individual's actions are strongly influenced by his desire to maintain a consistent cognitive condition with respect to his evaluations of himself. What this means is that when student ratings are much poorer than an instructor's self-ratings, a condition of imbalance (Heider, 1958), dissonance (Festinger, 1957), or incongruency (Newcomb, 1961; Secord & Backman, 1965) is created in the instructor. In an attempt to become more consistent, or in more theoretical terms to restore a condition of equilibrium, the instructor changes in the direction indicated by his students' ratings.

These theories assume, of course, that most instructors place enough value on collective student opinion, and that instructors know how to go about making changes. Undoubtedly some teachers merely write off student judgment as unreliable or unworthy, and for these individuals, changes are unlikely even though they may be called for. At least the changes are

unlikely if the only motivation comes from within the individual teacher. Increasingly, however, student ratings of professors are becoming public information, and in these instances there is undoubtedly a good deal of social pressure to change. In fact, not only is there social pressure, but in some instances there is economic pressure, since the ratings may be used in salary and tenure deliberations. But as I've said, it is not always clear to the teacher how to change, if indeed he or she believes the change would be an improvement. And this leads me to the implications of the second finding from my five-college study.

I mentioned that with additional time and with some interpretative information, the ratings for a more diverse group of teachers had changed in a positive direction. Not surprisingly, many teachers need more time to change their procedures, particularly in those areas that cannot be quickly altered (clarifying course objectives, for example). Yet if student ratings are to have <u>maximum</u> impact, I believe we need to do more in interpreting the results to instructors and in helping them improve. One of the reasons that we need to help instructors interpret their ratings is that the ratings are typically skewed in a positive direction. Most of us already know this, but the average teacher does not. On a five-point scale, he views his mean score of 3.6 as above average, when actually it may well be only average or even below average if compared to other teachers. Parenthetically, I might add that instructor self-ratings, not surprisingly, are skewed even more positively than student ratings. And faculty peer ratings based on classroom visits, according to some data I've recently collected, are also generally more favorable than student ratings. In any event, some kind of normative or comparative

data is important for interpreting student ratings, and, perhaps, the more the better. The instructor might be given the choice of comparing his students' responses to those of other teachers at his institution, or to those of members of his department; or perhaps he may prefer a more cosmopolitan comparison--such as to instructors from a sample of other institutions, or perhaps to a national sample of teachers in his field. The point is that a variety of comparisons might be made available to the instructor so that he can decide which are most meaningful.

Some of these comparison data are already being made available to instructors, though not always with the variety I've suggested. But I'm afraid that they do not totally solve the problem. There will still be some instructors who need special help, and for this reason Kenneth Eble (1971), for one, has suggested that individual instructional counseling be made freely available. A teacher counselor might not only help instructors interpret their student evaluations but could, of course, also suggest particular ways in which to improve. A few institutions are already doing this, but in these times of tight money this will probably remain a limited endeavor.

I'd like therefore to mention another possibility that I'm now pursuing. In place of an individual counselor I would propose substituting the next best thing: the computer. One of the remarkable feats of the computer is that it can be programmed to produce a verbal interpretation of a numerical summary. Rather than means, standard deviations, or percentile ranks, each professor could instead get several paragraphs of prose telling him how he differs from his own expectations and how he differs from some predesignated group, such as other teachers in his field.

The number-leery professor need not worry about whether his scores are significantly different--the computer will make that interpretation. Moreover it would even be possible to refer the instructor to specific materials, books, or even video tapes pertinent to his weaknesses. For example, if students said his course objectives were not made clear, or if they rated the quality of exams poorly, there would be several excellent references dealing with these topics suggested to the instructor. In fact, there's really no need to rely on the computer to produce these suggestions--we ought to be doing that sort of thing right now.

Before moving on to discussing other categories, I'd like to make one last point regarding the effects of student ratings on the individual teacher. With the emphasis generally put on mean scores or percentile ranks of scores, I'm afraid that the individual teacher is being influenced to see his class only as a homogeneous glob. Anyone who has taught knows that quite frequently there are several types of students in the typical class, each of which may be reacting a little differently to the teacher and the course. These different types and their various viewpoints do not mean that the ratings are unreliable in the sense that there is a great deal of fluctuation or inconsistency in student responses. We know that student ratings <u>are</u> reliable, as indicated by the numerous intraclass reliability studies that have been reported. What I'm talking about is identifying subgroups of students who differ systematically in their ratings. Is there, in short, some rhyme or reason to the diversity of viewpoints that may exist in the typical class?

One way to investigate this question is to use factor analytic techniques that allow one to group individuals rather than items as is usually

the case (see Tucker & Messick, 1963). The only study I have found that looked at this question had investigated students' general notions about types of teachers rather than their specific ratings of individual teachers (Rees, 1969). So I've undertaken some additional analyses--first with three large classes separately and then across a larger sample of courses--which indicate that there are frequently three or sometimes four points of view represented in a single class. Each of these groups sees various aspects of the course or the instruction they are receiving somewhat differently than the other groups. One group, for example, may have rated the instructor as generally ineffective, but at the same time indicated that the instructor was well organized and usually accessible; another group might have rated the instructor as ineffective and inaccessible. Unfortunately, I don't at this point have enough information about student characteristics that would allow me to describe the groups. Ultimately, however, it may be possible to alert the individual teacher to relevant subgroups or points of view in the class; these points of view might be identified by student characteristics information, or they might be identified by patterns of ratings. Until then, teachers should be encouraged to look at the distribution of student responses to the items on their rating form--and not only at the mean scores. While no one expects them to please all of their students all of the time, instructors ought to be aware of how they interact with different segments of the class.

Impact on Teaching Generally

Closely related to the effects of student ratings on the individual teacher is the possible impact that they have on teaching generally. The critics of student ratings claim that an undue emphasis on the ratings,

such as using them to assist in decisions on faculty promotions, can have adverse effects on instruction. What are some of these adverse effects? First, some critics claim that the ratings do not allow for individual styles of teaching, that they instead force everyone to be measured on the same yardstick. Few people would try to assess artists or composers on the same yardstick, according to one skeptic of student ratings. That skeptic goes on to say, in an article in The American Scholar, that:

> The art critic need not evaluate portraits painted by Picasso, Whistler, and Rembrandt in terms of criteria for effectiveness common to all three. He finds it possible to examine each artist's work in terms of the artists' own goals, or to identify the strengths and weaknesses of an individual painting in terms of relations of parts to the whole (Kossoff, 1972, p. 89).

Even though I don't happen to believe that teaching and art are entirely comparable, we know enough about teaching to know that individuals can have quite different styles, and that they should probably develop the style that best fits their personality and approach. I'll return to this point in a minute.

A second adverse effect of student ratings, according to the same critics, is that they encourage traditional modes of teaching. Most rating forms are indeed directed at classes taught in some combination of lecture-discussion, but logically so--that happens to be the way most courses have been taught and the forms are merely reflecting what is typically the case. The question is, however, are other methods such as student-centered learning, or nondirective teaching, or team teaching being stifled by the typical student rating forms? The answer, in my opinion, is that they are if an institution does not allow some flexibility in the application of student ratings. This means that for some courses, and this is still a relatively

small number on most campuses I suspect, it is necessary either to supplement or disregard items in the traditional rating forms.

Flexibility in the employment of student ratings is, in other words, extremely critical. Many of the widely used forms have been developed through what might be called the consensus approach. In other words the developers have asked samples of faculty members (or faculty members and students) to identify specific characteristics that are important in teaching. Those areas or items for which there was the greatest consensus were then included in the rating instrument. Generally speaking, the items have centered around such factors as course organization, teacher-student interaction, and communication or verbal fluency. It's clear that this approach does not produce an instrument that reflects any particular theory of teaching. And that probably has made good sense in view of the fact that it would be difficult to get any college faculty to agree on a single theory of teaching.

While most forms allow individual instructors to add their own items to a basic set, there are other ways in which the rating forms can be even more flexible. If the items are to be used in making decisions on faculty members, then the individual teacher might be allowed to eliminate those items that are not relevant to his style. Better yet, a system might be implemented which allows teachers to both choose and weigh in advance the items which they feel most adequately reflect their style of teaching and what they are trying to accomplish in the course. At least one institution is now working on such an approach.

Impact on Administrators

Another group that student ratings influence--albeit more indirectly than previous groups--are college administrators. I have two observations to offer regarding this. First, that in instances where the ratings are used in making decisions on promotions, it may well be that the dean or department chairman's job becomes a little easier.

National surveys have told us that frequently the judgments of one or more administrators are relied on to assess teaching effectiveness, particularly at smaller colleges. Not many people would defend this as a very wise or valid approach. If we can assume that the evidence provided by student evaluations means not only wiser decisions but also ones that are more easily defended, then students' evaluations make the administrators' jobs easier and more effective. Some, I realize, would debate that point.

A second observation that I have is that student evaluations may well be contributing to what seems to be a current groundswell for administrator evaluations by faculty members. A not too infrequent request to ETS is for an instrument to evaluate administrator performance. Apparently the feeling is that if faculty can be evaluated by their constituents, then by all means so can administrators. Increasingly, it would appear that they are. For example, the trustees of the State University of New York announced in January that the presidents of the 29 colleges operated by the state will have to undergo intensive evaluation of their records every five years. But I'm not at all sure that a handy-dandy machine-scored instrument could be developed that would measure reliably and validly an administrator's performance. More likely the charge is for

administrator accountability (to use the still-currently "in" word), in which an individual is accountable not only to his superiors but also to his subordinates.

Impact on Students

According to the results of the ACE 1972 annual survey of freshmen, students feel generally that faculty promotions ought to be based in part on student ratings. That opinion was endorsed by three-quarters of the students from the 373 institutions in the survey. This probably comes as no surprise. The past decade has, of course, been a time when students have demanded a greater role in institutional decision-making, and the evaluation of teaching would appear to be an area in which they feel they can make a unique contribution. Where student ratings have been incorporated into faculty evaluation procedures, therefore, the impact on students is likely to be quite positive; at least each of them can feel that he or she is helping the institution make important educational decisions. This is not to be taken lightly. While in the past teachers and administrators have been willing to give students a say in such areas as the establishment of student personnel policies and regulations, they've been more reluctant to relinquish their hold on academic decision-making.

Aside from this, probably the major impact of student ratings on students is provided by published course and teacher critiques. While some institutions make public the results of college-sponsored student evaluations (and some publish course guides based on detailed descriptions provided by the instructor), most of the critiques are based on surveys that are student initiated and conducted. As you might suspect, these student-produced critiques vary considerably in quality from one institution

to another; in fact, they may vary from year to year at single institutions, depending on which students get involved. The worst of the critiques have been based on poor samples and frequently border on sensationalism by highlighting the juiciest of criticisms. Needless to say these critiques do neither the teachers nor the students who purchase them much good. But what about the better publications; what about the critiques based on thorough methodology and which, as in some instances, also give the teacher an opportunity to respond to his student evaluations? Do they have a suitable reason for being? One might argue that they provide information that the college catalog or other publications don't provide and this would seem to be a valid purpose. Nevertheless there are many faculty members who object strongly to student conducted course ratings. Their objections have been delineated by Kerlinger in a 1971 article in School and Society. He argues that student initiated ratings result in "instructor hostility, resentment, and distrust," and thus alienate faculty members from their work. He goes on to suggest that ratings are legitimate only if conducted voluntarily by professors and used for self-improvement. Obviously then, not only is there concern for who initiates and conducts a student rating of instruction program, but also to what end the results are to be used.

Needed, it seems to me, is a major study of the effects of student ratings when they are used to assist in deciding whom to promote. There are a number of questions that such a study might investigate. For example, to what extent do faculty become alienated? Which types become most alienated? Does it encourage traditional teaching and limit teaching styles, as already discussed? Does it erroneously reinforce the notion in students that the instructor is largely responsible for how much students

learn in a course? This last point may be true regardless of how student rating results are used and in spite of the fact that many of the rating forms ask students about their own effort and involvement in the course. But the major question to be answered by such a study is whether more defensible promotion decisions are made when student evaluations are included as part of faculty assessment.

Impact on the College

The last category that I will comment on is the impact, or possible impact, of student ratings on the college.

I've already discussed changes that take place among individual teachers--or at least among some teachers. But can an institution, or perhaps t' departments within an institution, learn something about themselves from student evaluations? A corollary question is: "What can the institution or department then do about what they've learned?"

Let's start at the department level. A seldom mentioned though seemingly worthwhile use of student ratings is that of providing departments with information about the effectiveness of their offerings as seen by students. To do this it would be necessary to combine the ratings of all members in a department, and items dealing with specific as well as general course objectives should be included in the assessment. In addition to these course-instructor evaluations, a sort of major field questionnaire might be given to seniors. Princeton University, for one, has been using a major field or department questionnaire for the past several years. While not the typical application of student evaluations, the assessment of departmental offerings would seem to be worth consideration by other institutions.

Another point that might be made concerning the departments is that, as many of us have discovered, there are some interesting variations in the evaluations that teachers in different subject fields receive. Among a group of some 450 teachers, for example, I found that courses in the natural sciences, relative to those in humanities, social sciences, and education and applied subjects, were seen by students as having a faster pace, as being more difficult, and as being less likely to stimulate student interest. In addition, teachers perceived the natural science teachers in the sample as less open to other viewpoints. Humanities teachers, in comparison to those in the other three general subject areas, were less likely to inform students of how they were to be evaluated, and there was less agreement between the announced objectives of humanities courses and what was actually taught.

The obvious question is whether it is the subject matter itself that produces these differences or the types of individuals within each of the subject areas. It may well be a combination of both. At any rate, patterns of ratings would indicate that subject fields or departments might focus on certain apparent weaknesses (for example, humanities professors might attend workshops on improving their evaluation procedures).

The whole notion of focusing on weaknesses highlighted by student evaluations could be applied at the college level even more generally. If a college is able to compare itself to other colleges--that is, if the aggregate ratings of all teachers can be compared--then it may be possible to identify specific weaknesses. Workshops in that particular aspect of instruction might then be offered to assist in faculty improvement.

Conclusion

In this paper I've attempted to discuss the effects or possible effects of student evaluations on academia. It has been apparent throughout the discussion that the major effects are, to a large extent, dependent upon how the ratings are used. Their primary uses can perhaps be summarized best by adapting Michael Scriven's (1967) terms for the two major functions of tests: formative and summative evaluation. Tests used formatively, according to Scriven, give the instructor periodic feedback on his students' progress, thus telling the instructor what needs to be stressed in the future. The summative function of tests, as the term implies, is a way of providing a summative evaluation of each student at some point in time.

When student ratings of instruction are used formatively--that is, when they are used by instructors as a source of feedback on their teaching--the evidence indicates that some changes are made by the instructor. And most likely we can improve on this with better interpretation of the results. The effects of using student ratings in a summative way--that is, in making administrative decisions on faculty--is a little more difficult to assess. As a researcher I feel we ought to learn more about the side effects. But if I were a department chairman or dean faced with increasingly tougher tenure-promotion decisions, or if I were a faculty member who felt that his teaching was not being rewarded, then I might hold a different view. Certainly student evaluations are no less trustworthy than other methods now available to assess teaching performance, and when combined with other methods, they probably contribute to a fair judgment.

In closing, I'd like to return briefly to the title of this talk. As you have realized by this time, I don't believe that students, through student ratings, are or will become the Mario Puzo type of Godfather to the academic community. But this is not to say that they might not function in a limited way as proper Godfathers. Traditionally, of course, a Godfather has had a much more positive image; he essentially is one who helps provide guidance and direction to those in his charge. While I'm not suggesting that students are the new saviors of academia, or that college teachers must rely on the guidance of their students, I do think that a well-designed student ratings program can do more to benefit than to harm the academic community.

References

Centra, J. A. The utility of student ratings for instructional improvement. Project Report 72-16. Princeton, N. J., Educational Testing Service, 1972.

Eble, K. The recognition and evaluation of teaching. Project to improve college teaching, Salt Lake City, Utah, 1971.

Festinger, L. A theory of cognitive dissonance. Evanston, Ill.: Row, Peterson, 1957.

Heider, F. The psychology of interpersonal relationships. New York: Wiley, 1958.

Kerlinger, E. Student evaluation of university professors. School and Society, October 1972, 353-356.

Kossoff, E. Evaluating college professors by "scientific" methods. The American Scholar, Winter 1972, 79-93.

Newcomb, T. M. The acquaintance process. New York: Holt, Rinehart and Winston, 1961.

Rees, K. D. Dimensions of students' points of view in rating college teaching. Journal of Educational Psychology, 1969, 60(6), 476-482.

Scriven, M. The methodology of evaluation. American Educational Research Association monograph series on curriculum evaluation, No. 1, Perspectives of curriculum evaluation, 1967.

Secord, P. F., & Backman, C. W. An interpersonal approach to personality. In B. A. Maher (Ed.), Progress in experimental personality research, Vol 2. New York: Academic Press, 1965.

Tucker, L. R, & Messick, S. An individual difference model for multidimensional scaling. Psychometrika, 1963, 28(4), 333-367.